UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/714,331	11/14/2003	Pawan Goyal	ARC920030046US1	4042

7590 07/11/2007

Mark C. McCabe
IBM Corporation, Intellectual Property Law
650 Harry Road, Dept. C4TA/J2B
San Jose, CA 95120-6099

EXAMINER	
HOLLOWAY, DAVID A	
ART UNIT	PAPER NUMBER
2109	
MAIL DATE	DELIVERY MODE
07/11/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/714,331	GOYAL, PAWAN	
	Examiner	Art Unit	
	David A. Holloway	2109	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *14 November 2003*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-28* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☒ Claim(s) *24* is/are allowed.

6)☒ Claim(s) *1-23 and 25-28* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *14 November 2003* is/are: a)☐ accepted or b)☒ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *11142003*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

1. Claims 1-28 are pending in this application.

Specification

2. The disclosure is objected to because of the following informalities: misspelled word in the abstract, line 1. "finder grain" should be "finer grain". Appropriate correction is required.

Drawings

3. The drawings are objected to because of inconsistencies in Figure 1 and the specification. Item #32 contained within server two, #14 is not mentioned in the specification, and items #28, #30, and #32, should be labeled as #38, #40, and #42, respectively, in order to agree with the specification's description of Figure 1 found on page 10 of the specification. Figures 3A and 3B have misspelled words. "FO" should be "of". Corrected drawing sheets in compliance with 37 CFR 1.121(d) are required in reply to the Office action to avoid abandonment of the application. Any amended replacement drawing sheet should include all of the figures appearing on the immediate prior version of the sheet, even if only one figure is being amended. The figure or figure number of an amended drawing should not be labeled as "amended." If a drawing figure is to be canceled, the appropriate figure must be removed from the replacement sheet, and where necessary, the remaining figures must be renumbered and appropriate changes made to the brief description of the several views of the

drawings for consistency. Additional replacement sheets may be necessary to show the renumbering of the remaining figures. Each drawing sheet submitted after the filing date of an application must be labeled in the top margin as either "Replacement Sheet" or "New Sheet" pursuant to 37 CFR 1.121(d). If the changes are not accepted by the examiner, the applicant will be notified and informed of any required corrective action in the next Office action. The objection to the drawings will not be held in abeyance.

Claim Rejections - 35 USC § 112

4. The following is a quotation of the second paragraph of 35 U.S.C. 112: The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

5. Claims 1-6 and 15 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

6. As to claim 1, the phrase "for optimization of the multiple workloads across the multiple servers" on the last line of claim 1 is not clear. What is being optimized and how it is being optimized is not clear. It will be interpreted ast theoverall utilization of the server resources of the multiple servers that are executing the multiple workloads is what is being optimized.

7. As to claim 5, the phrase "maintaining an optimum utilization level across the plurality of servers" of claim 5, lines 3-4, is not clear. Again it is not clear exactly what is

being optimized and how it is being optimized. It could be th overall utilization of the resources across the multiple servers. However, there are multiple resources that are being managed, and it is not clear what the optimum utilization of the multiple would be. If the intent is to maintain the overall optimum utilization of the resources covered, i.e., CPU usage, memory resource usage, disk resource usage, and network bandwidth usage, then it is not clear how that differs from what is contained in claim 1 concerning the optimization of multiple workloads across the multiple servers. Nonetheless, it will be interpreted as "maintaining an optimum utilization level across the plurality of servers" indicates maintaining an optimum utilization of the collective server resources across the plurality of servers.

8. As to claim 15, the phrase "reassigning the VMs according changes in the identified resource requirements" on lines 2 and 3 of the claim is unclear. It will be interpreted as the amount of work assigned to each VM will be adjusted according to the identified resource requirements of the workloadassigned to each VM.

Claim Rejections - 35 USC § 103

9. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

10. Claims 1-13 are rejected under 35 U.S.C. 103(a) as being unpatentable

over Noland et al. (US 7,080,378), hereinafter Noland in view of Kawamoto et al. (US 7,117,499), hereinafter Kawamoto, in view of Sexton et al. (US 6,854,114), hereinafter Sexton, and further in view of Gootherts (US 2002/0099759 A1).

11. As to claim 1, Noland discloses a system to provide finer grain control in optimizing multiple workloads (Abstract, lines 1-3, here a workload is distributed across virtual servers, and the workload is considered to consist of multiple service requests. The multiple workloads are comprised of these multiple service requests), across multiple servers comprising a plurality of servers to be utilized by multiple workloads (Abstract, lines 1-3, the virtual servers are the multiple servers),.

12. Noland further discloses that each of the multiple workloads are distributed across the plurality of servers (Abstract, lines 6-8, last sentence, server requests are allocated among the cluster of virtual servers, where each virtual server corresponds to a physical server, see Fig.2, #202 - #205).
Noland further discloses that fractions of each of the multiple workloads are handled by separate entity in each server (the workload is distributed across multiple servers, and the workload will be spread across such entities of the servers).

13. Noland does not disclose a plurality of virtual machines (VMs) at each of the plurality of servers. However, Kawamoto discloses a plurality of virtual machines (VMs) in a virtual computer system (Abstract, lines 1-6, the logical partitions are the VMs).

14. Noland and Kawamoto are analogous art, because they are both in the

same field of endeavor of computer systems.

15. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him or her to incorporate the logical partitions, i.e., virtual machines of Kawamoto in the system of Noland. The motivation for doing so would have been to allow the system to control resource allocations that need to be shared between users accessing the computer system through the World Wide Web (Kawamoto, col. 1, lines 25-29).

16. Noland does not also disclose resource management logic to distribute server resources to each of the plurality of VMs according to predicted resource needs of each of the multiple workloads. However, Kawamoto discloses resource management logic to distribute server resources to each of the plurality of VMs (Abstract, last sentence, the adaptive control module comprises the resource management logic that determines the allocation of resources to the VMs) according to predicted resource needs of each of the multiple workloads (Fig. 6 shows that if adaptive control is enabled that new allocation ratios are determined, see Fig. 6, #2003 and #2004, thus, allocation resources needs are being predicted, also see col. 9, lines 29-41).

17. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him to incorporate the resource management logic that distributes the resources according to predicted resource needs as taught by Kawamoto in the system of Noland (Kawamoto, Abstract, last sentence, and col. 9, lines 29-41). The motivation for doing so would have

been to allow resources to be optimize the allocation of resources to the logical partitions, i.e., VMs.

18. Neither Noland nor Kawamoto necessarily disclose that the plurality of VMs at each of the plurality of servers each serve a different one of the multiple workloads (the term workload is being interpreted as being a load belonging to a given user as described in the specification). However, Sexton discloses that the plurality of VMs in a physical computer each serve a different one of the multiple workloads (Abstract, lines 1-2,), although, Sexton's system only includes a single physical computer.

19. Noland, Kawamoto, and Sexton are analogous art because they are all in the same field of endeavor of computer systems.

20. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto, and the teachings of Sexton before him to have the plurality of VMs at each of the servers to serve a different one of the multiple workloads. The motivation for doing so would have been to enable the operating system to ensure that the appropriate degree of insulation exists between the VM instances (Sexton, last sentence of Abstract).

21. Neither Noland, Kawamoto nor Sexton disclose that the fractions of each of the multiple workloads handled by each of the VMs can be dynamically adjusted to provide for optimization of the multiple workloads across the multiple servers. However, Gootherts discloses that the fractions of each of the multiple

workloads handled by each of the VMs can be dynamically adjusted to provide for optimization of the multiple workloads across the multiple servers (Abstract, lines 6-10, threads are transferred from the heaviest loaded server to the least loaded)

22. Noland, Kawamoto, Sexton, and Gootherts are analogous art because they are all in the same field of endeavor of computer systems.

23. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland, Kawamoto, Sexton, and the teachings of Gootherts before him or her to dynamically adjust the fractions of the workloads handled by each of the VMs to balance processing loads (Gootherts, Abstract, lines 1-2).

24. As to claim 2, Kawamoto discloses that the distribution of server resources to each of the plurality of VMs further comprises distributing server resources to the plurality of VMs according to the current and predicted resource needs of each of the multiple servers (col. 10, lines 47-51, the moving average of the loads is used to determine, i.e., predict the new allocation ratio, so the current load is used in predicting the new allocation ratio. Therefore, the new allocation is based on a predicted value which takes into account the current load).

25. As to claim 3, Kawamoto discloses that the server resources comprise percentage CPU, percentage of network bandwidth, disk resources and memory resources (Fig. 4, #1400, the server resources that can be used for measuring the load are CPU, network, disk, and memory usage, also see col. 7, lines 14-19. Although the user interface only allows one resource

allocation to be set at a time, Fig. 7 indicates that all 4 resources allocation types can be designated, see Fig.7, #2007, #2009, #2011, and #2013).

26. As to claim 4, Gootherts discloses that the finer grain control is achieved through recognizing when one of the plurality of servers is overloaded and shifting work to another of the plurality of servers which is not overloaded (Abstract, lines 6-10, threads are transferred from the heaviest loaded server to the least loaded).

27. As to claim 5, Gootherts discloses that the fractions of the multiple workloads being handled by the plurality of VMs (the VMs are present in the combined system of Noland, Kawamoto, and Sexton) can be dynamically adjusted in response to workload changes at the plurality of servers (the plurality of servers are also present in the combined system), wherein the dynamic adjustment provides for maintaining an optimum utilization level across the plurality of servers (Abstract, lines 6-10, threads are transferred from the heaviest loaded, source state server to the least loaded sink state server, where a source state server is a server where at least one thread is starving, i.e., waiting to execute, and a sink state server is a server that can accept additional threads without creating a starvation situation (col. 3, [0046], lines 3-5, and col. 3, [0047], lines 1-4). In claims 8-10 Gootherts describes criteria for choosing which server is the heaviest loaded and which is the least loaded. This criteria is based on a processor score that is at least the function of a processor state.

28. Gootherts does not explicitly state that the processor state includes the resources available to threads running on the processor. However, it would have been obvious to one skilled in the art that resource allocations would have an effect on the threads running on the processor. More specifically, starving threads will result from a lack of resources being available to a thread in a VM of the combined system of Noland, Kawamoto, Sexton and Gootherts. Although Gootherts only discloses transferring threads from one server to another server, in the combined system, the thread will be transferred from a VM on one server to a VM on another server).

29. As to claim 6, Kawamoto discloses that the optimum utilization level can be configured automatically via server management software or manually by a user with administrative privileges (Fig. 4 shows the input user interface. If adaptive control is enabled (Fig. 4, #1601 and #1602 radio buttons determine if the adaptive control is enabled) the resource allocations will be dynamically adjusted, see col. 6, lines 24-33. Otherwise the resource allocation will be statically set, see col. 9, lines 29-32 as well as Fig. 6, #2003).

24. As to claim 7, Noland discloses that the workloads are each distributed over a subset of the plurality of VMs (Abstract, lines 1-3, here a workload is distributed across virtual servers, and the workload is considered to consist of multiple service requests. The multiple workloads are comprised of these multiple service requests).

25. As to claim 8, Noland in view of Kawamoto, and Sexton, further in view of

Gootherts discloses that each VM in the subset of the plurality of VMs exists at a separate one of the plurality of servers as applied to claim 1 (Noland teaches distributing a workload over multiple servers, and Sexton teaches allocating one virtual machine per user session on a single server, so there are multiple servers with multiple VMs running on them, and each of the VMs belonging to a given user must exist on a separate server).

26. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland, Kawamoto, Sexton, and Gootherts before him to have each VM in the subset of the plurality of VMs exists at a separate one of the plurality of servers. The motivation for doing so would have been the same as the reasons given above in claim 1.

27. As to claim 9, Noland discloses that the workload distribution comprises distributing the work according to resources available to each of the VMs within the subset (Abstract, lines 12-15, workload is distributed among the cluster of virtual servers, and in view of Kawamoto each virtual server has multiple virtual machines running on it, so in the combined system, work will be assigned to a virtual machine on one of the virtual server in the cluster. Noland discloses a control program (col. 3, lines 33-36) that dispatches virtual machines from a list based upon parameters such as priority, I/O status, memory overhead support, i.e., resource status).

28. As to claim 10, Noland discloses that the system further comprises at least one global resource allocator to monitor resource distribution between the plurality of VMs (col. 1, lines 62-65, the invention functions by monitoring the existing performance of the virtual servers, and this is based on resources in the

virtual servers, such as CPU usage, see col. 1, lines 65-67).

29. As to claim 11, Noland discloses that the system further comprises at least one load balancer to measure the current offered load (col. 1, lines 65-67, col. 2, lines 1-2, the invention monitors resource loads and deploys more virtual servers if necessary).

30. As to claim 12, Noland discloses that the global resource allocator (col. 3, lines 47-40, the controller program is the global resource allocator) determines how to distribute the resources between the plurality of VMs (there are multiple VMs on each virtual server in the combined system of Noland and Kawamoto, and the controller program manages the resources used by the virtual servers (and on a lower level, the resources of the virtual machines), see col. 3, lines 47-49), according to the measurements received from the at least one load balancer (col. 1, lines 65-67, col. 2, lines 1-2 describes the load balancer, and it should be clear that if the system is managing the resources that are used by the virtual servers then it has to be based on information of the loads).

31. As to claim 13, Kawamoto discloses that each of the plurality of servers includes a local resource control agent (the hypervisor running on each server is the local resource control agent and receives resource allocation varying instructions from the adaptive control module), but Kawamoto does not disclose that each of the local resource control agents receive and implement instructions from the global resource allocator describing how the resources are to be distributed between the VMs located at each of the plurality of servers. However, in the combined system of Noland and Kawamoto, the adaptive control

module is running on the primary server, i.e., the server running the controller program, and the adaptive control module sends out commands to the hypervisors on each of the virtual servers to control the allocation of resources to each of the VMs.

31. Claims 14, 16-18, 20-23, 25, and 27 are rejected under 35 U.S.C. 103(a) as being unpatentable over Noland in view of Kawamoto.

32. As to claim 14, Noland discloses a server optimization device, for providing finer grain control in a virtual machine (VM) based hosting architecture, comprising at least one load balancer component to identify resource requirements for multiple different workloads in the VM based hosting architecture (col. 1, lines 65-67, col. 2, lines 1-2, the invention monitors resource loads and deploys more virtual servers if necessary), a global resource allocator partitioning component (the controller program) to assign VMs from multiple server machines to a workload according to the identified resource requirements (if the resource requirement is larger than what the allocated virtual servers can handle, a new virtual server is allocated and the workload will be sent there, so the workload is assigned to VMs according to resource requirements of the workload, also see col. 3, lines 33-36, virtual machines are dispatched from a list based upon parameters such as priority, I/O status, memory overhead support, i.e., resource status).

33. Noland does not disclose that the global resource allocator partitioning component to assign resources at each of the multiple server machines to the assigned VMs. However, Kawamoto discloses that the global resource allocator

partitioning component to assign resources at each of the multiple server machines to the assigned VMs according to the identified resource requirements (the adaptive control module is the global resource allocator partitioning component and it sends message to the hypervisor to allocate appropriate resources to the VMs according to the identified resource requirements)

34. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him to incorporate the resource allocating feature of Kawamoto in the system of Noland in order to optimize the performance of the individual servers.

35. As to claim 16, Noland does not disclose a plurality of resource allocator components at each of the multiple server machines, wherein the plurality of resource allocators are responsible for creating VMs and assigning VMs to workloads in response to instructions received from the global resource allocator partitioning component.

36. However, Kawamoto discloses a plurality of resource allocator components at each of the multiple server machines, wherein the plurality of resource allocators are responsible for creating VMs and assigning VMs to workloads (the resource allocators are the hypervisors and responsible for virtual machine management in general, including virtual machine creation and assignement of workloads to the virtual machines) in response to instructions received from the global resource allocator partitioning component (the adaptive control module is the global resource allocator partitioning component and sends commands to the hypervisor).

37. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him or her to have the hypervisors on each of the individual servers create VMs and assign VMs to workloads in response to instructions issued from the global resource allocator partitioning component. The motivation for doing so would have been to allow the controller program to manage the virtual machines (in the combined system of Noland and Kawamoto) similarly to the way the adaptive control module was sending messages to the hypervisor to manage the virtual machines (Kawamoto, col. 9, lines 37-39, here a resource varying instruction is issued from the adaptive control module to the hypervisor, but the hypervisor performs general virtual management not just resource allocation).

38. As to claim 17, Noland discloses that at least one load balancer continuously monitors the resource requirements for the multiple different workloads (col. 1, lines 62-65, the virtual server performance is monitored by the invention, ostensibly by the controller program, which performs the load balancing) and provides changes to the resource requirements of each of the multiple different workloads to the global resource allocator partitioning component (col. 1, lines 65-67, col. 2, lines 1-2, the controller program is the global resource allocator partition component but it also manages the load balancing, so it has the updated resource requirement information of the workloads).

39. As to claim 18, Noland discloses that the global resource allocator partitioning component issues instructions to the plurality of resource allocator components at each of the multiple server machines (col. 3, lines 47-50, the

control program manages the resources of the virtual servers, so there is implicity an entity on the virtual servers that allocate resources in response to control messages from the control program).

40. Noland further discloses that the redistribution of the server resources provides for optimizing workload across the multiple server to prevent the over-utilization or under-utilization of the multiple server machines (col. 3, lines 33-36, virtual machines (servers) are assigned to workload according to various resource parameters).

41. Noland does not disclose that the issued instructions provide for redistributing server resources to each of the VMs within each of the multiple server machines. However Kawamoto discloses that the issued instructions provide for redistributing server resources to VMs within a single machine, but Kawamoto does not disclose redistributing resources to VMs across multiple servers. In the combined system of Noland and Kawamoto there are multiple servers, so resources would be redistributed to VMs across multiple servers in the combined system.

42. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him to use the issued instructions to enable server resources to be redistributed optimally to the VMs (Kawamoto, Abstract, lines 1-4).

43. As to claim 20, Noland does not disclose that the resources comprise percentage of CPU, percentage of network bandwidth, disk resources and

memory resources. However, Kawamoto discloses that the resources comprise percentage of CPU, percentage of network bandwidth, disk resources and memory resources (Fig. 4, #1400, the server resources that can be used for measuring the load are CPU, network, disk, and memory usage, also see col. 7, lines 14-19. Although the user interface only allows one resource allocation to be set at a time, Fig. 7 indicates that all 4 resources allocation types can be designated, see Fig.7, #2007, #2009, #2011, and #2013).

44. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawomoto before him to enable the resources that can be redistributed to comprise percentage of CPU, percentage of network bandwidth, disk resources, and memory resources, since these resources are the basic resources required for program execution.

45. As to claim 21, Noland discloses that the instructions are being issued automatically via server management software in order to maintain a pre-defined level of optimization within the system. (the controller program balances the load to keep the virtual servers from being overloaded, and the controller program is server management software. There is no mention of the level of optimization being changed during operation, so it is obvious that the level of optimization is pre-defined)

46. As to claim 22, Noland discloses that the multiple different workloads are distributed over a subset of the assigned VMs (Abstract, lines 1-3, here a workload is distributed across virtual servers, and the workload is considered to consist of multiple service requests. The multiple workloads are comprised of these multiple service requests)

47. As to claim 23, Noland discloses that the multiple different workloads are each assigned to a customer application utilizing the VM based hosting architecture (it is clear that an application is running and the hosting architecture being used is VM, see col. 3, lines 31-33).

48. As to claim 25, Noland discloses a server optimization means, for providing finer grain control in a virtual machine (VM) based hosting architecture, comprising a means for identifying resource requirements for multiple different workloads in the VM based hosting architecture (col. 1, lines 65-67, col.2, lines 1-2, the invention monitors resource loads and deploys more virtual servers if necessary), a means for assigning VMs from multiple server machines to a workload according to the identified resource requirements (if the resource requirement is larger that what the allocated virtual servers can handle, a new virtual server is allocated and the workload will be sent there, so the workload is assigned to VMs according to resource requirements of the workload, also see col. 3, lines 33-36, virtual machines are dispatched from a list based upon parameters such as priority, I/O status, memory overhead support, i.e., resource status).

49. Noland does not disclose a means for assigning resources at each of the multiple server machines to the assigned VMs according to the identified resource requirements. However, Kawamoto discloses a means for assigning resources at each of the multiple server machines to the assigned VMs according to the identified resource requirements (the adaptive control module sends messages to the hypervisor to allocate appropriate resources to the VMs

according to the identified resource requirements).

50. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him to incorporate the resource allocate feature of Kawamoto in the system of Noland in order to optimize the performance of the individual servers.

51. As to claim 27, Noland does not disclose that the server optimization means further comprises a means for creating VMs and assigning VMs to workloads in response to instructions received from the global resource allocator partitioning component. However, Kawamoto discloses a means for creating VMs and assigning VMs to workloads in response to instructions received from the global resource allocator partitioning component (the resource allocators are the hypervisors, see col. 9, lines 37-39, and it is well known in the art that hypervisors are responsible for virtual machine management in general, including virtual machine creation and assignment of workloads to the virtual machines, see col. 2, lines 34-38) in response to instructions received from the global resource allocator partitioning component (the adaptive control module is the global resource allocator partitioning component and sends commands to the hypervisor)

52. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him to provide a means for creating VMs and assigning VMs to workloads in response to instructions received from the global resource allocator partitioning component in the system of Noland. The motivation for doing so would have

been to allow the controller program to manage the virtual machines (in the combined system of Noland and Kawamoto) similarly to the way the adaptive control module was sending messages to the hypervisor to manage the virtual machines (Kawamoto, col. 9, lines 37-39, here a resource varying instruction is issued from the adaptive control module to the hypervisor, but the hypervisor performs general virtual management not just resource allocation).

53. Claim 19 is rejected under 35 USC 103(a) as being unpatentable over Noland in view of Kawamoto, and further in view of Sexton.

54. As to claim 19, neither Noland nor Kawamoto disclose that the VMs at each of the multiple server machines serve a different one of the multiple different workloads. However, Sexton discloses that the VMs at each of the multiple server machines serve a different one of the multiple different workloads (Abstract, lines 1-2,), but Sexton's system only includes a single physical computer.

55. Noland, Kawamoto, and Sexton are analogous art because they are all in the same field of endeavor of computer systems.

56. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto, and the teachings of Sexton before him to have the plurality of VMs at each of the servers to serve a different one of the multiple workloads. The motivation for doing so would have been to enable the operating system to ensure that the appropriate degree of insulation exists between the VM instances (Sexton, last sentence of Abstract).

57. Claims 15, 26, and 28 are rejected under 35 USC 103(a) as being unpatentable over Noland in view of Kawamoto and further in view of Gootherts.

58. As to claim 15, neither Noland nor Kawamoto disclose that the global resource allocator partitioning component reassigns the VMs according to changes in the identified resource requirements. However, Gootherts discloses that threads are transferred from the heaviest loaded server to the least loaded server (Abstract, lines 6-10). In the combined system of Noland, Kawomoto, and Gootherts, the threads would be transferred from a VM on one server to a VM on another server according to changes in the identified resource requirements).

59. Noland, Kawamoto, and Gootherts are analogous art because they are all in the same field of endeavor of computer systems.

60. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto, and the teachings of Gootherts to reassigns the VMs according to changes in the identified resource requirements. The motivation would have been to optimize the load across the servers.

61. As to claim 26, neither Noland nor Kawamoto disclose that the optimization means further comprises a means for reassigning the VMs according to changes in the identified resource requirements. However, Gootherts discloses that the optimization means further comprises a means for reassigning the VMs according to changes in the identified resource

requirements (Abstract, lines 6-10, threads are transferred from the heaviest loaded server to the least loaded server). In the combined system of Noland, Kawomoto, and Gootherts, the threads would be transferred from a VM on one server to a VM on another server according to changes in the identified resource requirements.

62. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto, and the teachings of Gootherts to reassigns the VMs according to changes in the identified resource requirements. The motivation would have been to optimize the load across the servers.

63. As to claim 28, Noland discloses a computer program product for use with a computer hosting architecture, for providing finer grain control in a virtual machine (VM) based hosting architecture, comprising a computer-readable medium means, provided on the computer-readable medium, for identifying resource requirements for multiple different workloads in the VM based hosting architecture (col. 1, lines 65-67, col.2, lines 1-2, the invention monitors resource loads and deploys more virtual servers if necessary), a means for assigning VMs from multiple server machines to a workload according to the identified resource requirements (if the resource requirement is larger that what the allocated virtual servers can handle, a new virtual server is allocated and the workload will be sent there, so the workload is assigned to VMs according to resource requirements of the workload, also see col. 3, lines 33-36, virtual machines are dispatched from a list based upon parameters such as priority, I/O status, memory overhead support, i.e., resource status).

64. Noland does not disclose a means for assigning resources at each of the multiple server machines to the assigned VMs according to the identified resource requirements. However, Kawamoto discloses a means for assigning resources at each of the multiple server machines to the assigned VMs according to the identified resource requirements (the adaptive control module sends messages to the hypervisor to allocate appropriate resources to the VMs according to the identified resource requirements).

65. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto before him to incorporate the resource allocate feature of Kawamoto in the system of Noland in order to optimize the performance of the individual servers.

66. Neither Noland nor Kawamoto disclose means, provided on the computer-readable medium, for reassigning the VMs from multiple server machines to a workload according to changes in the identified resource requirements. However, Gootherts discloses means, provided on the computer-readable medium, for reassigning the VMs from multiple server machines to a workload according to changes in the identified resource requirements (Abstract, lines 6-10, threads are transferred from the heaviest loaded server to the least loaded server). In the combined system of Noland, Kawomoto, and Gootherts, the threads would be transferred from a VM on one server to a VM on another server according to changes in the identified resource requirements.

67. It would have been obvious to a person having ordinary skill in the art at the time of invention having the teachings of Noland and Kawamoto, and the

teachings of Gootherts to reassign the VMs from multiple server machines to a workload according to changes in the identified resource requirements. The motivation would have been to optimize the load across the servers.

Allowable Subject Matter

68. Claim 24 includes allowable subject matter. The following is a statement of reasons for the indication of allowable subject matter: The claim predicts what the load of a client application will be up front and then splits up work between multiple virtual machines, where each machine resides on a separate server. Based on the prediction of the load it also sets up resources on each of the virtual machines (which are on separate servers).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to David A. Holloway whose telephone number is (571)270-1899. The examiner can normally be reached on mon-fri 8:00 am - 5:00 pm (alternate fridays off).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Nabil El-Hady can be reached on (571)272-3963. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DH 7/6/2007

NABIL M. EL-HADY
SUPERVISORY PATENT EXAMINER